The Reference Class Problem for Credit Valuation in Science

Carole J. Lee (c3@uw.edu)

Abstract: Scholars belong to multiple communities of credit simultaneously. When these communities disagree about how much credit to assign to a scholarly achievement, this raises a puzzle for decision theory models of credit-seeking in science. The reference class problem for credit valuation in science is the problem of determining to which of an agent's communities – which reference class – credit determinations should be indexed for any given act under any given state of nature. I will identify strategies and desiderata for resolving ambiguity in credit valuation due to this problem and explain how pursuing its solution could, ironically, lead to its dissolution.

1. Introduction

Within the scientific community, there is a common understanding that its reward system drives problematic behavior linked to publication patterns, pipeline retention, hypercompetitive scientific cultures, and reproducibility. Conversely, there is also a shared sentiment that, in order to change these cultures and behaviors in ways that would improve science, the scientific community must coordinate across institutions to change how credit is assigned at the level of the individual scientist (Alberts et al. 2014, Nosek et al. 2015, Aalbersberg et al. 2017, National Academies of Sciences 2018, National Science Foundation 2015, Blank et al. 2017). The hope is

that increasing individual researchers' incentives towards increased transparency and openness will improve the integrity, reproducibility, and accuracy of the published record.[1]

Analogously, philosophers working in the "credit economy" tradition adopt the working assumption that there is some amount of credit that agents can accrue for different acts under different states of nature. This assumption allows them to use decision theory to model how credit-seeking among individual scientists can give rise to behavior and norms that support or thwart the achievement of community-wide goals. When, in the aggregate, individual credit-seeking cuts against collective ends, their approach can explore how changes to individuals' incentive structures can nudge and redirect individual behavior (Bruner and O'Connor 2017, Rubin and O'Connor 2018, Bright 2017, Heesen 2017, Kitcher 1990, Strevens 2003, Zollman 2018). Different philosophers make different assumptions about the norms by which credit gets allotted – for example, whether credit is best thought of as all-or-nothing (Strevens 2003, Bright 2017, Heesen 2017) or as something that may come in degrees (Bruner and O'Connor 2017, Rubin and O'Connor 2018, Zollman 2018). However, the general approach assumes that there is some precise way to assign credit to different acts under different states of nature – an assumption that allows these philosophers to model credit-seeking behavior and the emergence of scientific norms in formally tractable ways.

But, how much credit gets assigned to any given act under any given state of nature? Just as each of us simultaneously belongs to multiple social categories each of which is tied to implied social hierarchies (Macrae, Bodenhausen, and Milne 1995, Crenshaw 1989), each

[1] Institutions can also experience incentives that promote or thwart scientific ends (Lee and Moher 2017).

scholar simultaneously belongs to multiple communities of value with implied social hierarchies for assigning credit. To which of an agent's communities – which reference class – should credit determinations be indexed and why?

In this paper, I will use examples from the current context of science's complex and dynamic culture to motivate and illuminate what I will call the *reference class problem for credit valuation in science*. I will identify a few strategies and desiderata for solving ambiguity in credit assignments due to the reference class problem. And, I will say a bit about how developing the resources needed to solve it could ultimately sow the seeds for its own dissolution.

2. The Reference Class Problem for Credit Valuation in Science

The contours of this puzzle about the "coin of recognition" (Merton 1968, 56) become visible when one moves beyond thinking about credit in generic, abstractions of scientific communities towards the heterogeneous communities we find today. I start from this slightly more concrete perspective because prestige requires recognition *by individuals and forums* that are themselves valued by credit-seeking scholars (Zuckerman and Merton 1971, Lee 2013): credit worthiness in science is a function of the individuals and systems designed to assess, allocate, dispute, and enforce it. Although some aspects of Zuckerman and Merton's narrative about the origins of the normative structure of science have been contested by historians (Csiszar 2015, Biagioli 2002), we see the social dynamics Zuckerman and Merton proposed clearly at play in contemporary science. For example, Nature Publishing Group recently found that – for the 18,354 authors in science, engineering, and medicine surveyed – the reputation of a journal is the primary factor driving choices about where to submit their work, where reputation is

primarily determined by the journal's impact factor and whether it is "seen as the place to publish the best research" (Nature Publishing Group 2015). Factors associated with a journal's ability to archive and disseminate research – things like a journal's time from acceptance to publication, indexing services, or Open Access options – were much less important.[2]

Within academia, each of us simultaneously belongs to multiple communities of value. The reference class problem arises when these different communities of value disagree about the amount of credit an agent accrues for choosing some act under some state of nature. Although I take this problem is be general, for the sake of clarity and simplicity in presentation, I will focus my examples on communities that can be described as having a nesting structure: for example, individual scholars belong to specific sub-disciplines, which are nested within disciplines, which are nested within a more general population of scholars. A sub-population that is nested within a population can have a credit sub-culture whose valuations differ from that of the population, whose valuations can differ from that of the super-population. In these cases, changing how narrowly or broadly one draws the boundaries of an agent's community of valuation can change the amount of credit assigned to a scholarly accomplishment. This gives rise to the *reference class problem for credit valuation in science*: to which of the agent's communities – which reference class – should credit valuations be indexed when determining the amount of credit the agent accrues for different acts under different states of nature?

[2] I recognize that some decision theorists, especially those working outside of philosophy, may reject or remain agnostic about attributing mental states such as beliefs to agents (Okasha 2016). However, because I understand credit and credit-seeking as sociological phenomena involving status beliefs such as these, I am committed to attributing beliefs to agents.

There are many examples across academia where nesting community structures can give rise to paradoxes and pathologies in credit assignments. For example, scholars' individual sense of what counts as quality work – their individual credit assignments – may deviate from what is endorsed in a sub-discipline or discipline's status hierarchy (Correll et al. 2017, Centola, Willer, and Macy 2005, Willer, Kuwabara, and Macy 2009). A puzzle that has cachet in a sub-discipline may be of peripheral importance within that discipline: for example, a more accurate technique for measuring how temperature cools with elevation considered critical in mountain meteorology and mountain ecology (Mindner, Mote, and Lundquist 2010) may have less visibility, despite its relevance, to the larger discipline of hydrology (Livneh et al. 2013). A question or technique that is thought to have high impact across fields (e.g., machine learning) may have little prominence within some of those fields.

Hypothetically speaking, one could imagine differences in valuations giving rise to a *Simpson's paradox in credit valuation*. Simpson's paradox is a phenomenon whereby a trend that appears in a population reverses or disappears when it is disaggregated into sub-populations (Blyth 1972). For example, a classic study found that, when looking at aggregate graduate school admissions data at UC Berkeley, women were, on the whole, less likely than men to be accepted; however, when the data was disaggregated into admitting departments, women were more likely than men to be admitted (Bickel, Hammel, and O'Connell 1975). Analogously, a *Simpson's paradox in credit valuation in science* would occur in cases where a population-level preference for scholarly product *a* versus *b* reverses when the population is disaggregated into its component sub-populations. In Simpson's Paradox cases, thinking more carefully about the context of evaluation usually leads to using a reference class that is finer-grained than the population-level. However, it's not clear whether this would always be the case in evaluations of

scientific credit. Hypothetically speaking, consider a hypothetical scenario in which an interdisciplinary project is not preferred by the individual disciplines represented by its authors or content, but is preferred when those disciplines are aggregated together. And, imagine that this project gets published in a journal, valued by those disciplines, that seeks papers of interest *across and beyond disciplines* (not just within disciplines): this is one way to interpret, for example, *Science's* mission to publish papers that "merit recognition by the wider scientific community and general public. . . beyond that provided by specialty journals" (Science). Which reference class would be most relevant in evaluating the value of this project?

There are other ways of dividing scholarly communities into nesting structures that create tensions in credit assignments. The pressures a scholar may feel from the incentive structure impacting her department/school may be slightly different from the incentive structure impacting her university. A coarse but concrete way to see this is to think about the prestige structure reified and reinforced by ranking systems (Espeland and Sauder 2012, 2016, Sauder and Espeland 2006), which transform "the ways professional opportunities are distributed" (Espeland and Sauder 2016, 7). An untenured business school professor with a potentially high impact manuscript needs to burnish her prestige in the eyes of both her dean and her provost, since both will evaluate her tenure case. If her provost is working to gain stature on the Academic Rankings of World Universities [ARWU], the professor should submit her manuscript to *Science* or *Nature*, since the ARWU ranks universities by their publications in these journals (Academic Ranking of World Universities 2018). However, if her dean is trying to gain stature on the *Financial Times* International ranking of MBA programs, she should submit to one of the fifty business, economics, or psychology journals by which the FT ranking system evaluates Business

school prestige – notably, the journal list does not include *Science* or *Nature* (Ormans 2016). What should the business school professor do?

Finally, credit assignments can vary depending on how long a time window a scholar keeps in view. A coarse but concrete way to think about this is by looking at how metrics for evaluating scholarship change over time. Journal impact factors are becoming less useful measures for evaluating an individual's scholarly contribution: since the advent of the digital age, the most elite journals (including *Science* and *Nature*) are publishing a decreasing percentage of the top cited papers (Larivière, Lozano, and Gingras 2013); the relationship between journal impact factor and paper citations has declined over time (Lozano, Larivière, and Gingras 2012); and, the citation distributions between journals "overlap extensively" (Larivière et al. 2016). The current wisdom is that if quantitative indicators are to be used to evaluate research, it is more useful to use article-level metrics such as citations as well as alternative metrics such as downloads and views (San Francisco Declaration on Research Assessment 2013, Hicks and Wouters 2015, Wilsdon et al. 2017). On the horizon, there are now calls for creating new metrics that can encourage researchers and journals to be transparent and open in their reporting practices (National Academies of Sciences 2018, Wilsdon et al. 2017, Aalbersberg et al. 2017). Note that, the rise of such metrics – as well as the growing meta-research literature that ranks journals by the replicability (Schimmack 2015) or sample size and statistical power of their published results (Fraley and Vazire 2014) – makes it possible for a journal's impact factor and epistemic credibility to come apart (Fang and Casadevall 2011).

Decision theorists capture the risky nature of individual choices by allowing for uncertainty about which states of the world will come to be; and, when the probabilities attached to different outcomes are understood subjectively, these models permit a kind of subjectivity in

estimates of expected credit for different acts. However, I hope the examples throughout this section animate genuine *ambiguity in credit* due to the reference class problem for credit valuation in science.

3. Strategies and Desiderata for Solving the Reference Class Problem

How might decision theorists try to solve the reference class problem for assigning credit in science? One possible approach argues for the "correctness" of using one community rather than another. For example, it might be tempting to argue that all prestige is discipline-based since many scholarly prizes are distributed for excellence in particular disciplines (e.g., Nobel prize, Fields prize, academic society prizes); and, even when research is funded or published in interdisciplinary contexts, it may be primarily evaluated on the basis of its disciplinary excellence (Lamont 2009, but see Lee et al. 2013). Indexing credit valuation to a particular community need not prevent scholars from outside that community from understanding the relative value of that contribution: for example, if one were to adopt the old-fashioned and problematic assumption that an article's impact can be measured by the impact factor of the journal in which it is published,[3] and one recognizes that citations rates vary across disciplines, one could use field-normalized percentiles to understand a paper's impact in a metric that is legible across fields (Hicks and Wouters 2015). Because this strategy for addressing the

[3] The citation distributions within journals are so skewed that it is statistically improper to infer the impact of an individual article on the basis of the impact factor of the journal in which it is published (San Francisco Declaration on Research Assessment 2013, Hicks and Wouters 2015, Wilsdon et al. 2017, Larivière et al. 2016, Wilsdon et al. 2015).

reference class problem relies heavily on identifying the "right" community, defending the centrality of the chosen community as opposed to others is critical. For example, some may challenge the idea that disciplines should be the sole arbiter of credit: note that the awarding of some scientific prizes reach across disciplinary conceptions of excellence (e.g., consider winners of the MacArthur Genius Prize and the psychologists who have won the Nobel Prize in Economics).

Another possible approach creates an algorithm that calculates the credit value of a scholarly contribution by summing the credit valuation of multiple communities. This approach would need to identify exactly how much to weight each community's valuation – with a rationale for why – since different weightings could lead to different overall credit valuations.[4] Note that some scholars take this style of approach when trying to measure the relative prestige of journals: in particular, the Eigenfactor score rates journals according to the number of its incoming citations, where the "relative importance" of each incoming citation is contextualized by the frequency with which the citing journal is itself cited (West, Bergstrom, and Bergstrom 2010).

Those who may wish to model the implications of different approaches for solving the reference class problem may try to do so by setting up hypothetical communities that assign

[4] On the face of it, this may seem like a form of commensuration because it involves summing values to calculate an overall score (Espeland and Stevens 1998). However, the process of commensuration requires combining values across *qualitatively* different domains of value. For clearer examples of commensuration in scholarly evaluation, see Lee (2015).

community boundaries and credit assignments in *de facto* ways to see what kinds of behaviors and norms emerge.

However, to solve the underlying conceptual problem, one must provide theories of community and credit that address two fundamental but vexing questions. How should one define and gerrymander the boundaries of the relevant communities invoked in the proposed solution? And, how does one determine the amount of credit those communities would assign to different acts under different states of nature? These questions may not be independently answerable. The boundaries of a community may need to be defined in terms of patterns of shared lore among its members about how credit is accrued – shared beliefs that coordinate credit-seeking and enforcement behavior in cases where status beliefs are internalized as norms (Merton 1973) and in cases where they are not (Willer, Kuwabara, and Macy 2009, Ridgeway and Correll 2006). Conversely, in recognition that some community members can have more influence than others on the content of reigning status beliefs, a community's credit assignments may need to be defined with some reference to the causal patterns of interaction among specific individuals and clusters of individuals – including status judges who wield "social control through their evaluation of role-performance and their allocation of rewards for that performance" (Zuckerman and Merton 1971, 66). Note, however, that answers to these questions should not *exclusively* inform each other. Notably, we must be careful not allow the size of a scholarly population and/or the power of its status judges to fully determine the intellectual value of the questions pursued by any particular partition of the scholarly universe.

4. Conclusion

Scientific credit – the "coin of recognition" (Merton 1968, 56) – is assessed, allocated, disputed, and enforced by many different communities and institutions within science that support and sustain a multiplicity of status hierarchies. This gives rise to what I have called the reference class problem for credit valuation in science. Solving this problem requires developing rich theories of community and credit that are based on fine-grained information about the structure and status systems of complex scholarly networks. The irony of this assessment is that such investigation towards solving the reference class problem could ultimately sow the seeds for its own dissolution.

In particular, such study can render friable a critical assumption for both the reference class problem and for decision theory models: namely, that communities, once defined, assign determinate amounts of monistic credit for different acts under different states of nature – that credit "can vary quantitatively but not qualitatively" (Anderson 1993, xii).[5] Contrary to this, recent policy papers call for moving away from narrowly conceived measurements of research excellence towards broader ones that are sensitive to the diversity of individual researchers', programs', and academic institutions' research missions (Hicks and Wouters 2015, Wilsdon et al. 2015). Such work can include community-engaged scholarship that creates, disseminates, and implements knowledge in coordination with the public to identify social interventions, change social practice, and influence policy (Hicks and Wouters 2015, San Francisco Declaration on Research Assessment 2013, Boyer 1990, Escrigas et al. 2014). From the

[5] Note too that, for formal reasons, the assumption that individual credit assessments could be aggregated into a collective one is questionable given the challenges of combining individual preferences into collective ones (Arrow 1950).

perspective of these efforts, plurality in our notions of scholarly excellence and credit – and differences in valuation and prioritization practices between individuals and communities – may be best conceived, not as a logical problem to solve, but as a starting point for theorizing.

Acknowledgments: Many thanks to Christopher Adolph, Aileen Fyfe, Crystal Hall, Jessica Lundquist, Conor Mayo-Wilson, and Kevin Zollman for helpful conversations. This research used statistical consulting resources provided by the Center for Statistics and the Social Sciences, University of Washington.

References

Aalbersberg, IJsbrand Jan, Tom Appleyard, Sarah Brookhart, Todd Carpenter, Michael Clarke, Stephen Curry, Josh Dahl, Alex DeHaven, Eric Eich, Maryrose Franko, Len Freedman, Chris Graf, Sean Grant, Brooks Hanson, Heather Joseph, Véronique Kiermer, Bianca Kramer, Alan Kraut, Roshan Kumar Karn, Carole Lee, Aki MacFarlane, Maryann Martone, Evan Mayo-Wilson, Marcia McNutt, Meredith McPhail, David Mellor, David Moher, Alison Mudditt Mudditt, Brian Nosek, Belinda Orland, Tim Parker, Mark Parsons, Mark Patterson, Solange Santos, Carolyn Shore, Dan Simons, Bobbie Spellman, Jeff Spies, Matt Spitzer, Victoria Stodden, Sowmya Swaminathan, Deborah Sweet, Anne Tsui, and Simine Vazire. 2017. "Making science transparent by default; Introducing the TOP Statement." *OSF Preprints*. doi: https://doi.org/10.31219/osf.io/sm78t.

Academic Ranking of World Universities. 2018. "ShanghaiRanking's Academic Ranking of World Universities 2018 Press Release." accessed September 1.

http://www.shanghairanking.com/Academic-Ranking-of-World-Universities-2018-Press-Release.html.

Alberts, Bruce, Marc W. Kirschner, Shirley Tilghman, and Harold Varmus. 2014. "Rescuing US biomedical research from its systematic flaws." *Proceedings of the National Academy of Sciences* 111 (16):5773-7.

Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Arrow, Kenneth J. 1950. "A difficulty in the concept of social welfare." *Journal of Political Economy* 58 (4):328-46.

Biagioli, Mario. 2002. "From Book Censorship to Academic Peer Review." *Emergences: Journal for the Study of Media & Composite Cultures* 12 (1):11-45.

Bickel, P. J., E. A. Hammel, and J. W. O'Connell. 1975. "Sex bias in graduate admissions: Data from Berkeley." *Science* 187 (4175):398-404.

Blank, Rebecca, Ronald J. Daniels, Gary Gilliland, Amy Gutmann, Samuel Hawgood, Freeman A. Hrabowski, Martha E. Pollack, Vincent Price, L. Rafael Reif, and Mark S. Schlissel. 2017. "A new data effort to inform career choices in biomedicine." *Science* 358 (6369):1388-9.

Blyth, Colin R. 1972. "On Simpson's Paradox and the sure-thing principle." *Journal of the American Statistical Association* 67 (338):364-66.

Boyer, Ernest L. 1990. *Scholarship Reconsidered*. San Francisco, CA: The Carnegie Foundation for the Advancement of Teaching.

Bright, Liam Kofi. 2017. "On Fraud." *Philosophical Studies* 174:291-310.

Bruner, Justin, and Cailin O'Connor. 2017. "Power, Bargaining, and Collaboration." In *Scientific Collaboration and Collective Knowledge*, edited by Thomas Boyer-Kassem, Conor Mayo-Wilson and Michael Weisberg, 135-157. Oxford, UK: Oxford University Press.

Centola, Damon, Robb Willer, and Michael Macy. 2005. "The emporer's dilemma: A computational model of self-enforcing norms." *American Journal of Sociology* 110 (4):1009-40.

Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, and Sandra Nakagawa. 2017. "It's the conventional thought that counts: How third-order inference produces status advantage." *American Sociological Review* 82 (2):297-327.

Crenshaw, Kimberle. 1989. "Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics." *University of Chicago Legal Forum* 139:139-168.

Csiszar, Alex. 2015. "Objectivities in Print." In *Objectivity in Science: New Perspectives from Science and Technology Studies*, edited by Flavia Padovani, Alan Richardson and Jonathan Y. Tsou, 145-69. Cham, Switzerland: Springer International Publishing.

Escrigas, Cristina, Jesús Granados Sánchez, Budd Hall, and Rajesh Tandon. 2014. "Editor's introduction. Knowledge, engagement and higher education: Contributing to social change." In *Report: Higher Education in the World*, edited by Cristina Escrigas, Jesús Granados Sánchez, Budd Hall and Rajesh Tandon. Palgrave Macmillan.

Espeland, Wendy Nelson, and Michael Sauder. 2012. "The Dynamism of Indicators." In *Governance by Indicators: Global Power through Quantification and Rankings*, edited by Kevin Davis, Angelina Fisher, Benedict Kingsbury and Sally Engle Merry, 86-109. Oxford: Oxford University Press.

Espeland, Wendy Nelson, and Michael Sauder. 2016. *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York, NY: Russell Sage Foundation.

Espeland, Wendy Nelson, and Mitchell L. Stevens. 1998. "Commensuration as a Social Process." *Annual Review of Sociology* 24:313-43.

Fang, Ferric C., and Arturo Casadevall. 2011. "Retracted Science and the Retraction Index." *Infection and Immunity* 79 (10):3855-9.

Fraley, R. Chris, and Simine Vazire. 2014. "The N-Pact Factor: Evaluating the Quality of Empirical Journals with Respect to Sample Size and Statistical Power." *PLOS ONE* 9 (10):e109019. doi: 10.1371/journal.pone.0109019.

Heesen, Remco. 2017. "Communism and the Incentive to Share in Science." *Philosophy of Science* 84:698-716.

Hicks, Diana, and Paul Wouters. 2015. "The Leiden manifesto for research metrics." *Nature* 520:429-31.

Kitcher, Philip. 1990. "The Division of Cognitive Labor." *The Journal of Philosophy* LXXXVII (1):5-22.

Lamont, Michèle. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

Larivière, Vincent, Véronique Kiermar, Catriona J. MacCallum, Marcia McNutt, Mark Patterson, Bernd Pulverer, Sowmya Swaminathan, Stuart Taylor, and Stephen Curry. 2016. "A simple proposal for the publication of journal citation distributions." *BioRxiv*:062109.

Larivière, Vincent, George A. Lozano, and Yves Gingras. 2013. "Are elite journals declining?" *Journal of the Association for Information Science and Technology* 65 (4):649-55.

Lee, Carole J. 2013. "The limited effectiveness of prestige as an intervention on the health of medical journal publications." *Episteme* 10 (4):387-402.

Lee, Carole J. 2015. "Commensuration bias in peer review." *Philosophy of Science* 82:1272-83.

Lee, Carole J., and David Moher. 2017. "Promote Scientific Integrity via Journal Peer Review." *Science* 357 (6348):256-7.

Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. 2013. "Bias in peer review." *Journal of the American Society for Information Science and Technology* 64 (1):2-17.

Livneh, Ben, Eric A. Rosenberg, Chiyu Lin, Bart Nijssen, Vimal Mishra, Kostas M. Andreadis, Edwin P. Maurer, and Dennis P. Lettenmaier. 2013. "A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions." *Journal of Climate* 26 (23):9384-9392.

Lozano, George A., Vincent Larivière, and Yves Gingras. 2012. "The weakening relationship between the Impact Factor and papers' citations in the digital age." *Journal of the American Society for Information Science and Technology* 63 (11):2140-45.

Macrae, C. Neil, Galen V. Bodenhausen, and Alan B. Milne. 1995. "The Dissection of Selection in Person Perception: Inhibitory Processes in Social Stereotyping." *Journal of Personality and Social Psychology* 69 (3):397-407.

Merton, Robert K. 1968. "The matthew effect in science." *Science* 1968:56-63.

Merton, Robert K. 1973. "The normative structure of science." In *The Sociology of Science: Theoretical and Empirical Investigations*, edited by Norman W. Storer, 267-78. Chicago, IL: University of Chicago Press.

Mindner, Justin R., Philip W. Mote, and Jessica D. Lundquist. 2010. "Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains." *Journal of Geophysical Research: Atmospheres* 115. doi: https://doi.org/10.1029/2009JD013493.

National Academies of Sciences, Engineering, and Medicine,. 2018. Open Science by Design: Realizing a Vision for 21st Century Research. Washington, D.C.: The National Academies Press.

National Science Foundation. 2015. Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science. In *Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences*.

Nature Publishing Group. 2015. "Author Insights 2015 Survey."

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Mahlhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni. 2015. "Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility." *Science* 348 (6242):1422-5. doi: 10.1126/science.aab2374.

Okasha, Samir. 2016. "On the interpretation of decision theory." *Economics & Philosophy* 32 (3):409-33.

Ormans, Laurent. 2016. "50 Journals used in FT research." accessed September 1. https://www.ft.com/content/3405a512-5cbb-11e1-8f1f-00144feabdc0.

Ridgeway, Cecilia L., and Shelley J. Correll. 2006. "Consensus and the creation and status beliefs." *Social Forces* 85 (1):431-53.

Rubin, Hannah, and Cailin O'Connor. 2018. "Discrimination and Collaboration in Science." *Philosophy of Science* 85:380-402.

San Francisco Declaration on Research Assessment. 2013. "The San Francisco Declaration on Research Assessment (DORA)." accessed September 1. https://sfdora.org/read/.

Sauder, Michael, and Wendy Nelson Espeland. 2006. "Strength in numbers? The advantages of multiple rankings." *Indiana Law Journal* 81 (1):205-27.

Schimmack, Ulrich. 2015. "Replicability Ranking of 26 Psychology Journals." January 18. https://replicationindex.wordpress.com/2015/08/13/replicability-ranking-of-26-psychology-journals/.

Science. "Mission and Scope." accessed September 1. http://sciencemag.org/about/mission-and-scope.

Strevens, Michael. 2003. "The role of the priority rule in science." *Journal of Philosophy* 100 (2):55-79.

West, Jevin D., Theodore C. Bergstrom, and Carl T. Bergstrom. 2010. "The Eigenfactor Metrics™: A network approach to assessing scholarly journals." *College & Research Libraries* 71 (3):236-44.

Willer, Robb, Ko Kuwabara, and Michael W. Macy. 2009. "The False Enforcement of Unpopular Norms." *American Journal of Sociology* 115 (2):451-90.

Wilsdon, James, Liz Allen, Eleonora Belfiore, Philip Campbell, Stephen Curry, Steven Hill, Richard Jones, Roger Kain, Simon Kerridge, Mike Thelwall, Jane Tinkler, Ian Viney, Paul Wouters, Jude Hill, and Ben Johnson. 2015. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management.

Wilsdon, James, Judit Bar-Ilan, Robert Frodeman, Elisabeth Lex, Isabella Peters, and Paul Wouters. 2017. Next-generation metrics: Responsible metrics and evaluation for open science. Report of the European Commission Expert Group on Altmetrics. European Commission.

Zollman, Kevin J. S. 2018. "The Credit Economy and the Economic Rationality of Science." *The Journal of Philosophy* 115:5-33.

Zuckerman, Harriet, and Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System." *Minerva* 9 (1):66-100.